信息化时代新闻素材挖掘技术的应用研究

李云霞 李 伟
（怀安县融媒体中心，河北 张家口 076150）



摘要：【目的】信息化时代的到来对新闻行业的发展具有积极促进作用。新闻行业已经逐渐从传统新闻的采集模式转变成了互联网时代的新闻采集模式。在这一背景下，如何快速准确地获取到所需要的新闻素材已经成为当前新闻行业所面临的一个重要问题。【方法】文章结合信息化时代背景，分析新闻素材挖掘技术在新闻行业中的应用场景与应用优势，剖析新闻素材挖掘技术的关键技术，从信息获取、素材处理、媒体监测、舆情分析和事件报道五个方面论述新闻素材挖掘技术的具体应用。【结果】利用相关技术来对海量信息进行挖掘处理，快速准确地获取所需要的新闻素材，提升工作效率。【结论】随着技术的不断进步和创新，新闻素材挖掘技术将会有更多的应用场景得到开发和实现，有效促进我国媒体行业更加健康快速的发展。

关键词：信息时代；新闻素材；挖掘技术；应用研究；新闻数据　　中图分类号：G2　　文献标识码：A
文章编号：1671-0134（2023）05-064-04　　DOI：10.19483/j.cnki.11-4653/n.2023.05.014
本文著录格式：李云霞，李伟．信息化时代新闻素材挖掘技术的应用研究 [J]. 中国传媒科技，2023（05）：64-67.

随着社会的不断发展，媒体行业也随之不断地发展。在新媒体环境下，传统的媒体形式已经逐渐被取代，人们获取信息的渠道越来越多，面临的信息选择也越来越多。这就对媒体行业提出了更高要求。如何在海量的信息中快速准确地获取到自己需要的信息成为当前媒体行业所面临的重要问题。因此，为了快速准确地获取到所需要的新闻素材，就需要利用相关技术来对海量信息进行处理。

1. 新闻素材挖掘技术在新闻行业中的应用场景与应用优势

随着信息化时代的不断发展，互联网已经成为人们获取信息最主要的途径之一。新闻素材是在某一特定领域中所具有的、能够反映该领域内的客观事实以及发展状况的所有文字、声音、图像等资料总和。新闻素材挖掘技术是指将这些新闻素材进行挖掘处理，并最终提取出能够反映当前社会现状及意义、价值的重要信息。[1]

1.1 应用场景

1.1.1 从海量的信息中发现具有新闻价值的内容

随着网络技术的快速发展，信息的传播变得越来越快，新闻事件的传播速度也越来越快。为了能够使新闻报道快速、全面、准确地传播，需要从海量的新闻中进行筛选和提炼，挖掘出具有新闻价值的素材。传统媒体在对新闻事件进行报道时，需要编辑人员对其进行整理和筛选，并从中找出有新闻价值的内容。而数据挖掘技术可以帮助编辑人员找到有新闻价值的内容，并将其提取出来，为媒体提供具有新闻价值的素材。

1.1.2 挖掘出那些已经被其他媒体报道过的内容

从各种媒体报道中寻找具有新闻价值的内容，不仅可以帮助记者发现一些未被报道过的消息和事件，还可以帮助记者发现一些已经被其他媒体报道过，但并没有被记者关注到的相关事件和人物。

1.1.3 帮助分析采访对象

在采访过程中，记者会遇到各种各样的人和事。通过对采访对象进行分析，可以为记者提供一个具有价值的采访对象。

1.2 优势

新闻素材挖掘技术在新闻报道中具有重要的应用优势，可以帮助新闻从业者更好地把握事件脉搏，提高工作效率和精准度。[2] 具体表现在：（1）通过该技术，实现对各种媒体平台上发布的新闻信息进行全面获取，包括文字、图片、视频等形式，不受时间和空间限制。（2）利用自然语言处理和机器学习等技术，快速对大量的新闻素材进行分析和分类，提高工作效率。（3）通过对新闻素材的挖掘和分析，追踪事件的发展历程，及时了解事件的最新进展和各方态度。（4）帮助新闻从业者对新闻事件进行深入的分析，准确把握事件背后的原因和影响因素。（5）通过分析新闻素材的特定要素，发现具有新闻价值的信息，包括热点话题、社会热议、人物故事等，对新闻报道的创新和发展有着积极的作用。（6）降低出错率。

由于新闻素材挖掘技术可以快速地对大量信息进行筛选和分类，从而减少新闻从业者的疏漏并降低其出错率。

2. 信息化时代新闻素材挖掘技术的关键技术

2.1 新闻素材的预处理技术

在信息化时代背景下，传统的新闻信息采集模式已经逐渐无法满足新闻行业发展的需要。因此，要提升新闻采集质量和效率，就需要应用相关的新闻素材挖掘技术，对新闻信息进行预处理。预处理技术是指在信息采集之前对所需要的信息进行数据转换、数据压缩以及数据抽取等操作。[3] 这一过程，主要是通过对采集到的大量数据信息进行整理和筛选，形成一个符合用户需求的数据集合。预处理技术的应用不仅可以提升新闻素材挖掘质量和效率，同时也能够提升新闻采集的质量和效率。此外，通过对大量数据信息进行有效的预处理，还能保证所获得信息的全面性和准确性，为后续新闻素材挖掘工作提供坚实的基础。

2.2 关键词提取技术

在传统新闻的采集过程中，需要大量的人力采集新闻信息，会出现大量的人工编辑、整理工作。这样不仅会耗费大量的时间，同时还会导致新闻内容存在较大的重复性。随着信息技术和新闻行业的不断发展，需要采集的新闻素材量也逐渐增多。对这些大体量的素材来说往往很难找到准确的关键词，导致无法将相关信息有效地提取出来。现在利用关键词提取技术则可以帮助工作人员快速获取到所需要的信息，同时还可以保证相关信息的完整性和准确性，为后续工作的开展奠定良好的基础。

关键词提取技术是通过相关软件自动抽取所需新闻内的关键词，并将这些关键词与新闻素材相匹配，从而完成对相关新闻的自动筛选。[4] 关键词提取技术主要包含关键词检测系统、信息管理系统等多个方面的内容。关键词检测系统主要是通过计算机软件对所需要获取的新闻素材进行自动检测，并通过计算机软件对这些新闻素材进行分类，与数据库中的相关信息进行匹配，从数据库中获取到所需要的信息。信息管理系统是通过计算机软件来存储、管理和控制相关数据，将所需要获取的数据和新闻素材进行自动分类整理。

总的来说，关键词提取技术可以有效提高新闻信息检索、新闻内容发现等多方面的效率和质量，对新闻素材的挖掘和新闻采编工作具有积极意义。

2.3 文本分类技术

在挖掘新闻素材时，需要对采集到的信息进行预处理，从采集到的原始数据中提取信息，对信息进行分类，在分类的过程中，需要使用到文本分类技术。

文本分类技术是将采集到的信息进行初步归类，将信息按照一定的标准划分为不同的类别。文本分类技术是目前新闻素材挖掘工作中应用较为广泛的一项技术。在使用过程中需要注意两点：一是需要确保分类信息没有重复；二是需要确保所划分出的类别具有较高的准确性。

2.4 情感分析技术

情感分析技术主要是通过提取新闻中所包含的情感信息，再通过情感分析技术对这些情感信息进行进一步的筛选，为新闻工作者提供有价值的新闻素材。

在应用情感分析技术时，要对文本进行预处理，主要是利用分词系统对文本信息进行去噪处理和分词处理，使其可以被计算机识别。在对文本信息进行预处理时，应该遵循一定的原则。首先确保所提取的情感信息是真实存在的情感，其次应确保所提取出来的情感信息是没有经过修改或者篡改过的情感信息。[5] 在这一过程中需要注意不能提取与新闻事件无关或者与新闻事件相关度不高的情感信息。

情感分析技术在新闻素材挖掘中有着广泛的应用，主要体现在：（1）舆情分析：识别和分类新闻素材中的情感倾向，通过对情感极性的判断和分析，对新闻事件的舆情进行监测和评估，帮助媒体了解公众对事件的态度和情感倾向。（2）事件检测与跟踪：帮助自动发现和跟踪新闻素材中的事件，并判断该事件从事者的情感倾向。[6] 通过监测和分析事件相关的情感信息，能更好地了解事件的发展趋势和关键点。（3）新闻推荐：根据用户的兴趣和情感偏好，智能推荐新闻素材，从而提高用户的阅读体验。例如，对积极情感偏好的用户，系统可以推荐一些正面的、鼓舞人心的新闻素材。（4）品牌监测：对品牌相关的新闻素材进行自动分析和监测，了解公众对品牌的情感倾向和态度，并且识别出可能对品牌形象产生负面影响的信息，帮助企业及时做出应对措施。

2.5 基于知识图谱的新闻素材挖掘技术

随着互联网技术和大数据技术等相关信息技术的不断发展，知识图谱在当前社会各个行业和领域中都得到了较为广泛的应用。知识图谱是指在一定的领域中，梳理和展示实体以及实体之间的关系，并对实体之间所具有的特征进行描述，最终使这些实体之间形成一个互相联系的网络。知识图谱中包含的信息都是真实有效的，并且与现实生活存在一定的联系，具有较强的实用性。[7]

基于知识图谱的新闻素材挖掘技术，是一种基于大数据和人工智能技术的新闻素材处理方法。它利用

自然语言处理、机器学习、图谱构建等技术，将新闻素材中的信息进行提取和分析，并将这些信息组织成为一张知识图谱，实现对新闻事件的深度挖掘和理解。

基于知识图谱的新闻素材挖掘技术主要有以下几个特点：（1）知识图谱构建：基于知识图谱的新闻素材挖掘技术，通过构建一个海量的知识图谱，将新闻素材中的各类信息进行关联和整合，从而实现对新闻事件的知识获取和理解。（2）语义分析：该技术利用自然语言处理技术，对新闻素材进行分析和处理，实现对文本中的语义信息的提取和理解，同时也能够识别出文本中的情感和态度。（3）事件发现：基于知识图谱的新闻素材挖掘技术能够自动发现新闻素材中的事件和话题，并对相关信息进行分类和整合，从而实现对事件的深度挖掘和理解。（4）智能推荐：该技术能够根据用户的兴趣和需求，自动推荐相关的新闻素材和事件，从而提高用户的信息获取效率。

基于知识图谱的新闻素材挖掘技术，将会在未来的新闻素材处理领域中具有广泛的应用前景，为新闻媒体和用户提供更加丰富、准确和深度的信息服务。

3. 新闻素材挖掘技术在新闻采写中的应用

3.1 新闻信息获取：通过新闻素材挖掘技术更快捷、准确地获取消息来源

在信息化时代，人们获取新闻素材的方式更加多样化，除了传统的搜索引擎，人们还可以通过社交媒体和网络论坛来获取信息。对新闻工作者来说，可以通过新闻素材挖掘技术，更快捷、准确地获取消息来源。

第一，可以通过新闻搜索引擎来获取信息。新闻搜索引擎是一种以新闻为主要检索对象的搜索引擎，通过对网络上广泛存在的新闻信息进行采集、整理和分类，提供给用户一个高效、准确的新闻检索工具。[8]例如谷歌新闻、百度新闻等，可以根据新闻的关键词、时间、地点、来源等多种方式进行检索。新闻搜索引擎不仅能够提供较为全面的新闻信息，而且搜索速度非常快。第二，社交媒体是一种基于用户生成内容的平台，它是获取实时新闻的重要渠道之一。例如推特、微博等，这些平台上的用户可以随时发布新闻、评论、照片、视频等内容，可在其中寻找有价值的新闻信息。通过关注新闻领域中的专家、媒体机构和记者，获得更加深入、全面和及时的新闻信息。第三，数据库检索。新闻数据库是一种专门存储新闻信息的数据库，其中包括历史性的新闻事件及最新的新闻报道。利用数据库检索技术，可以根据关键词、时间、地点、媒体等多种属性进行检索，获取具有一定权威性的新闻信息。第四，可以通过 RSS 订阅。RSS 是一种网上内容分发技术，它提供一种简单的方式来订阅新闻、博客和其他网站的更新。通过 RSS 订阅，用户可以选择自己感兴趣的新闻来源，每当发布新内容时，就会发送通知给用户。这对于经常需要查询某些新闻的人来说非常方便，同时也能够保证获取到最新的新闻信息。

通过以上四种方法，可以在短时间内获取到大量的新闻信息，通过新闻素材挖掘技术的帮助，进一步筛选、分类、分析这些信息，为新闻报道提供更加准确、全面、深入的素材。

3.2 新闻素材处理：通过新闻素材挖掘技术处理大量文本、筛选，提取出有价值信息

在大数据时代背景下，新闻采写工作人员要深入研究大数据技术。由于互联网上信息量较大，且内容繁杂、分散。所以在进行数据信息处理时要充分利用数据挖掘技术对数据信息进行筛选、整理、整合等操作。

首先需要对采集到的新闻文本进行清洗，去除无用的字符、符号、标点等，并做好格式规范化，以便后续处理。接着，采用自然语言处理技术，将新闻文本进行分词，即将连续的文本划分成一个一个单独的词语，以便后续处理；通过自然语言处理技术（如 TF-IDF 算法、TextRank 算法等），从分词后的文本中提取出关键词，以便后续分析和挖掘；通过主题分析或话题模型技术，对新闻文本进行分类，将其归入不同的主题或话题中，便于后续处理；利用机器学习等技术，对新闻文本中的情感进行判断和分类，识别出积极、消极或中性的情感倾向；通过命名实体识别技术，识别出新闻文本中的人物、地点、机构等实体，便于后续分析；通过关系提取技术，识别出新闻文本中不同实体之间的关联关系，便于后续分析。另外，在进行文本数据处理时，要根据不同的需求对其进行分类和筛选。如通过关键词过滤、文本聚类等方式来快速、准确地筛选出相关信息。

3.3 媒体监测：通过新闻素材挖掘技术监测媒体平台上的新闻信息，把握事件脉搏

新闻素材挖掘技术需要用到一个非常庞大的数据库，包含了大量的新闻素材。在实际工作中，编辑要合理利用新闻素材挖掘技术，要对媒体平台上发布的新闻进行全面监测，在此基础上结合自身经验对相关信息进行整理和分析，为后续的写作提供必要的资料和素材。

首先，通过对大量新闻文本进行挖掘和分析，实时监测和跟踪某一事件的发展情况，了解事件的起因、过程、结果等，以及与事件相关的各种信息，如主要参与者、时间、地点、背景等。这可以帮助新闻工作者及时了解事件的进展，从而更好地做出决策和行动。其次，通过分析新闻中的情感倾向、态度和评价等信息，

可以进行舆情分析和监测。[9] 例如，可以分析某一品牌或公众人物的声誉和形象在社会中的影响，及时发现和解决公众关注度高、评价多的问题。最后，通过对各大媒体平台的新闻报道进行监测，了解不同媒体之间的关注焦点、报道角度、数据来源等，以及对某个事件的态度和立场等。这可以帮助企业或公众人物更好地了解媒体的运作规律和特点，制定相应的传播策略，提升公众认知度和形象。

3.4 舆情分析：通过新闻素材挖掘技术分析公众的情感和反应，把握社会舆论态势

大数据时代下，新闻素材挖掘技术不仅能帮助编辑对新闻事件进行深入分析，还能对公众的情感和反应进行分析，进而帮助编辑更好地把握社会舆论态势。具体而言，新闻素材挖掘技术的应用可对公众的情感和反应进行分析，根据不同新闻事件中公众的情感和反应不同，进而对其进行分析、总结和归纳。例如，在日常生活中，当人们遇到一些令人感到沮丧或者感到气愤的事件时，他们会有各种不同的情感表现。在这种情况下，新闻采写人员可以利用新闻素材挖掘技术对公众的情感进行分析、总结和归纳。同时，在实际新闻采写过程中，新闻采写人员要掌握一定的分析技巧和方法,并不断学习和探索各种分析技巧和方法，更好地挖掘公众情感和反应。

3.5 事件报道：利用新闻素材挖掘技术收集、整理和分析事件信息，锁定更有深度、广度的报道素材

在大数据时代，数据是新闻事件的重要来源，具有一定的权威性。通过对数据的搜集、整理和分析，可以发现事件背后的隐性信息，为编辑提供更有深度和广度的报道素材。

首先，利用大数据技术对新闻事件进行搜集，可以发现事件中存在的隐性信息。[10] 如对事件中人员的年龄、性别等进行统计分析，发现该事件发生的特殊群体。此外，还可以分析该事件发生时出现的时间、地点等信息，以更好地理解事件发生背后所存在的意义和价值。其次，利用大数据技术对新闻事件进行整理和分析，发现该事件在社会中引起的反响，以及人们对这一问题的看法。通过对事件本身进行深度挖掘，发现该新闻事件产生的原因，以及该问题在社会上引起巨大反响后人们对其态度等信息。通过分析这些数据信息可以为编辑提供更有深度、广度和高度的报道素材，使编辑能够更好地把握社会发展趋势。

4. 信息化时代新闻素材挖掘技术的发展趋势与未来展望

随着信息技术的不断发展，新闻行业的发展不仅对当前社会具有很大的影响，同时也对我国未来的发展具有积极的促进作用。因此，在当前社会中，新闻行业为了能够更好地适应社会的发展需求，需要不断地创新新闻素材挖掘技术，不断提高新闻行业中所应用到的技术水平。

随着大数据时代的到来，信息数据呈现出爆炸式增长，新闻素材挖掘技术将这些海量的数据进行分析和筛选，为新闻媒体提供更加准确、全面和丰富的新闻素材。该技术最大的创新点是能够根据用户的需求进行大数据分析，并对用户的浏览行为、收藏内容、行为轨迹等进行综合分析，为用户提供个性化服务。在未来发展中可以应用于更多领域，例如城市建设、交通运输等领域，将其与人们的生活相结合，使人们获得更加方便快捷的服务。

参考文献

[1] 刘绍琴 . 广播电视新闻编辑技巧及提升策略 [J]. 中国报业，2022（10）：104-105.

[2] 张建华 . 互联网背景下新闻素材的深度开发和改良探析 [J]. 中国地市报人，2021（9）：34-36.

[3] 唐雪莲 . 融媒体视阈下电视新闻素材的整理与编辑 [J]. 采写编，2021（5）：44-45.

[4] 李轩冰 . 基于文本挖掘技术的网络热点新闻系统的研建 [D]. 北京：北京林业大学，2020.

[5] 孟伟 . 新媒体时代信息传播方式研究解析 [J]. 环球首映，2019（7）：155.

[6] 李心舒 . 基于 Spark 的多领域网络新闻热点挖掘技术研究与实现 [D]. 北京：北京邮电大学，2019.

[7] 冯霞 . 面向网络新闻的热点话题挖掘技术研究 [D]. 南京：南京理工大学，2019.

[8] 钟智锦，王童辰 . 大数据文本挖掘技术在新闻传播学科的应用 [J]. 当代传播，2018（5）：12-18.

[9] 罗政 .TRS 网络舆情和社会媒体挖掘服务平台 [J]. 中国传媒科技，2011（8）：54-56.

[10] 尚恩洁 . 没落“贵族”《东方时空》的机遇——浅谈电视新闻杂志利用数据挖掘技术发展 [J]. 新闻研究导刊，2015（14）：237+240.

作者简介：李云霞（1975-），女，河北张家口，编辑，研究方向为新闻采编；李伟（1983-），男，河北张家口，助理编辑，研究方向为新闻采编。

（责任编辑：张晓婧）